DIRECTORATUL NAȚIONAL
DE SECURITATE CIBERNETICĂ

DEEPFAKE

MANIPULAT SAU INFORMAT?

Notă: acest ghid a fost parțial elaborat cu ajutorul inteligenței artificiale (IA).

TLP:CLEAR

Publicul țintă al ghidului

Acest ghid se adresează persoanelor cu vârsta peste 18 ani, în vederea creșterii gradului de conștientizare și înțelegere asupra utilizării instrumentelor de inteligență artificială (IA) de tip generativ pentru crearea de conținut video, audio sau de imagini, ori privind detectarea utilizării unor astfel de instrumente.

Deepfake, o definiție

Deepfake este o manipulare digitală a unei înregistrări video, audio sau a unei imagini, realizată cu ajutorul inteligenței artificiale (IA) sau a altor programe specializate.

Un exemplu: Există multiple cazuri în care un videoclip cu o personalitate publică oferă sfaturi financiare. În acest videoclip, persoana recomandă insistent investiția într-o anumită companie sau afacere, promițând profituri mari și riscuri minime.

Detaliile contează: Videoclipul este realizat profesional, cu o calitate a imaginii și sunetului ridicată. Persoana publică pare sinceră și convingătoare, folosind un limbaj accesibil și argumente persuasive.

Scopul campaniei: Un astfel de Deepfake ar putea fi folosit pentru a manipula populația să investească într-o afacere ilicită (scam[1]), deținută sau controlată de manipulatori.

Impact: Videoclipul este distribuit rapid pe rețelele de socializare, cu un potențial uriaș de a fi vizionat de milioane de oameni. Mulți dintre cei care văd videoclipul ar putea fi convinși să investească în afacerea recomandată, riscând ulterior să piardă sume semnificative de bani.

De ce este important să înțelegem Deepfake

Înțelegerea Deepfake-urilor este crucială, având implicații semnificative în societate, politică, securitate cibernetică și în manipularea încrederii publice.

Impact asupra adevărului și încrederii: Deepfake-urile pot distorsiona realitatea, creând conținut video sau audio extrem de realist, care poate fi greu de distins de materialele autentice. Aceasta poate submina încrederea în mass-media și în informațiile pe care ne bazăm pentru a lua decizii informate.

Securitate și fraude: Tehnologia Deepfake poate fi folosită pentru a crea înregistrări false care implică persoane în activități pe care nu le-au realizat sau pentru a produce dovezi false în contexte juridice. De asemenea, poate fi utilizată în tentative de fraudă, cum ar fi falsificarea identității în apeluri telefonice sau video pentru a obține acces neautorizat la informații sau resurse financiare.

Manipulare și dezinformare: În politică și în alte domenii, Deepfake-urile pot fi folosite pentru a manipula opinia publică, a discredita adversarii sau a crea confuzie. Acestea pot influența alegerile, relațiile internaționale și pot alimenta teorii ale conspirației.

Impact social și etic: Utilizarea imorală a tehnologiei Deepfake, pentru crearea de materiale compromițătoare sau hărțuirea online, ridică probleme serioase de etică și respectare a drepturilor individuale.

Pregătire pentru viitor: Pe măsură ce tehnologia continuă să avanseze, este probabil ca Deepfake-urile să devină tot mai sofisticate și mai greu de detectat. Înțelegerea modului în care funcționează și a modalităților de a le identifica este esențială pentru dezvoltarea de tehnologii și strategii care să contracareze utilizarea lor rău intenționată.

Educație și conștientizare: Educația publicului despre existența și capacitatea Deepfake-urilor poate ajuta la atenuarea impactului lor prin creșterea scepticismului sănătos față de conținutul dubios și încurajarea verificării informațiilor din surse multiple.

Pentru toate aceste motive, este important să dezvoltăm un nivel colectiv de înțelegere a Deepfake-urilor și a potențialului lor de a afecta societatea. Acest lucru ne va permite să navigăm într-o lume digitală tot mai complexă cu discernământ și o mai mare precauție.

[1] Mod general de a defini diferite tipuri de fraudă cu care se confruntă utilizatorii atunci când folosesc Internetul. Cei care desfășoară astfel de activități folosesc instrumente tehnice sau tehnici de inginerie socială cu intenția de a obține foloase financiare.

Tehnologia din spatele Deepfake

Deepfake-urile sunt create folosind o combinație de tehnici de IA și învățare automată (Machine Learning - ML). Tehnologiile cheie implicate sunt:

1. **Rețele neuronale convoluționale (CNNs):** Sunt tipuri de rețele neuronale artificiale specializate în analiza imaginilor și a videoclipurilor. Ele sunt antrenate pe seturi mari de date, imagini și videoclipuri reale pentru a învăța caracteristicile faciale, expresiile, mișcările corpului și alte detalii vizuale.
2. **Rețele neuronale generative (GANs):** Rețele neuronale artificiale care pot genera conținut nou, realist, similar cu datele pe care au fost antrenate. În contextul Deepfake, GAN-urile sunt utilizate pentru a genera imagini și videoclipuri false care sunt foarte asemănătoare cu cele reale.
3. **Învățarea automată (ML):** Este utilizată pentru a antrena algoritmii Deepfake să identifice și să manipuleze elemente specifice ale imaginilor și videoclipurilor, cum ar fi expresiile faciale, mișcările buzelor, sincronizarea audio, etc.

Procesul de creare a unui Deepfake

Colectarea datelor: Pentru a crea un Deepfake convingător, este necesară o cantitate mare de date pentru antrenament care să includă imagini și videoclipuri capturate din diferite unghiuri și în diverse situații. Cu cât datele sunt mai variate și de calitate mai bună, cu atât Deepfake-ul va fi mai realist.

Antrenarea modelului: Utilizând datele colectate, algoritmi de IA și ML sunt antrenați pentru a identifica și învăța caracteristicile unice ale persoanei țintă, cum ar fi trăsăturile faciale, expresiile și modul în care se mișcă sau vorbește. Scopul acestei faze este de a permite modelului să reproducă aceste detalii cu o precizie cât mai mare.

Generarea Deepfake: După ce modelul este suficient antrenat, este folosit pentru a crea conținut falsificat, în care persoana țintă spune sau întreprinde acțiuni neadevărate. Etapa implică generarea de imagini sau secvențe video artificiale, dar extrem de realiste, utilizând capabilitățile modelului antrenat.

Manipularea detaliilor: Pentru a spori autenticitatea și credibilitatea Deepfake-ului, detaliile fine sunt ajustate meticulos. Aceasta include sincronizarea mișcărilor buzelor cu conținutul audio falsificat, corectarea expresiilor faciale pentru a se potrivi cu contextul generat și finisarea altor detalii minore care contribuie la realismul general al conținutului.

Distribuirea: Deepfake-ul final poate fi distribuit online prin intermediul rețelelor sociale, al platformelor de sharing video sau chiar prin mesagerie. Intenția distribuirii poate fi de a înșela, de a discredita o persoană sau de a manipula opinia publică.

Exemple reale de utilizare a Deepfake

Banca Națională a României (BNR) a avertizat publicul cu privire la o schemă de înșelăciune care implică utilizarea tehnologiei Deepfake pentru a crea videoclipuri false cu guvernatorul BNR. În aceste videoclipuri, guvernatorul pare să promoveze o platformă de investiții, însă BNR a declarat că acestea sunt false. Scamul folosește IA pentru a modifica vocea și imaginea guvernatorului, cu scopul de a induce în eroare publicul pentru a participa la investiții frauduloase, promițând câștiguri financiare rapide și ușoare.

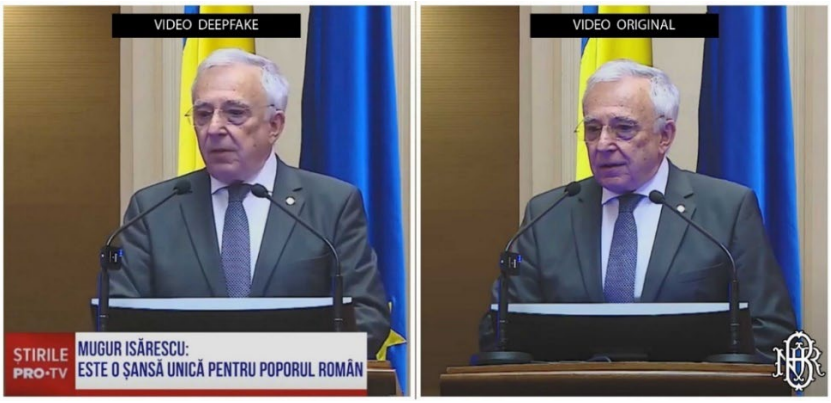


Figura 1 SCAM | Deepfake cu guvernatorul BNR Mugur Isărescu care promovează o platformă de investiții

O femeie pensionară din Vaslui a fost victima unei escrocherii postată online pe platforma YouTube. Escrocii au creat un videoclip fals în care un cunoscut bancher și alte personalități cunoscute recomandau o platformă de investiții. Promisiunea unor profituri rapide a convins pensionara să investească suma de 52.000 de lei, economii strânse în 20 de ani de muncă. Chiar dacă femeia a raportat această fraudă autorităților, experții consideră că șansele de recuperare a fondurilor pierdute sunt minime[i].

Actorii rău intenționați devin tot mai inventivi. Acum pot imita perfect vocea unei persoane dragi pentru a vă păcăli. Te pot suna pretinzând că sunt un membru al familiei care are nevoie urgentă de bani.

Recomandăm ca întotdeauna să fie verificată identitatea apelantului, chiar dacă pare să fie o rudă sau un prieten cunoscut. Nu trimite niciodată bani în grabă și anunță imediat autoritățile dacă ai suspiciuni. Doar prin vigilență și informare te poți proteja de această formă de fraudă[ii].

În 2019, un Deepfake audio a fost folosit pentru a înșela un CEO cu 220.000 de euro. Directorul general al unei firme de energie cu sediul în Marea Britanie a crezut că vorbea la telefon cu directorul general al companiei-mamă germane atunci când a urmat ordinele de a transfera imediat 220.000 de euro în contul bancar al unui furnizor maghiar. De fapt, vocea aparținea unui escroc care folosea tehnologia vocală IA pentru a-l imita pe directorul executiv german[iii].

Pericolul Deepfake în contextul electoral

În era digitală actuală, în care granița dintre realitate și ficțiune se estompează în mod constant, procesul electoral depășește simpla confruntare a ideologiilor și promisiunilor politice, transformându-se într-un teren complex de luptă ideologică. Tehnologiile Deepfake, capabile a sintetiza realist imagini și voci , pot influența semnificativ opinia și votul alegătorilor în timpul campaniilor electorale, având un impact major asupra procesului democratic, din următoarele considerente:

Impactul asupra politicienilor: Reputația unui politician poate fi grav afectată de videoclipuri Deepfake fabricate, care pot discredita imaginea sa și pot deteriora șansele de a câștiga alegerile. Răspândirea dezinformării prin Deepfake poate manipula percepția publică asupra caracterului și programului politici-anului, afectând negativ cariera sa politică.

Impactul asupra partidelor politice: Deepfake poate fi utilizat ca instrument strategic pentru a discredita partidele rivale, prin crearea de materiale false care să le prezinte într-o postură negativă. Promovarea agendei propriului partid prin Deepfake poate fi o modalitate eficientă de a influența opinia publică și de a atrage alegători. Utilizarea necorespunzătoare a Deepfake de către partidele politice poate duce la scandaluri și la pierderea încrederii publicului în sistemul politic.

Impactul grupurilor de interese: Deepfake poate fi utilizat de grupuri de interese din interiorul unei țări pentru a manipula opinia publică legată de probleme civice. Crearea de conținut falsificat adaptat la contextele locale poate intensifica sau calma nemulțumirile legate de anumite probleme sau evenimente. Utilizarea Deepfake de către grupurile de interese poate duce la polarizarea societății și la o erodare a discursului public democratic.

Tehnicile IA/Deepfake utilizate în influențarea campaniilor electorale:

- Falsificarea video realistă: Algoritmi avansați de IA creează videoclipuri aproape imposibil de distins de cele reale, prezentând scene false cu politicieni sau evenimente inventate.
- Clonarea vocii și falsuri audio: Vocea unui politician este reprodusă cu o precizie uimitoare, generând mesaje false atribuite în mod eronat persoanei respective.
- Generarea de text sintetic: Texte credibile sunt create imitând stilul și tonul unei personalități politice sau a unei instituții, răspândind dezinformare.
- Schimbarea fețelor și transformarea: Tehnologia permite modificarea fețelor în videoclipuri, creând scenarii false care nu s-au petrecut niciodată sau nu în contextul prezentat.
- Predicția comportamentală: IA analizează comportamentul online pentru a prezice răspunsul la anumite mesaje, facilitând crearea de Deepfake-uri țintite pentru a influența segmente specifice de alegători.

Aceste tehnici pot fi utilizate de actori malițioși pentru a discredita candidați, a manipula opinia publică și a submina încrederea în procesul electoral. Este esențială conștientizarea pericolelor Deepfake și implementarea de măsuri pentru a combate dezinformarea și proteja integritatea alegerilor.

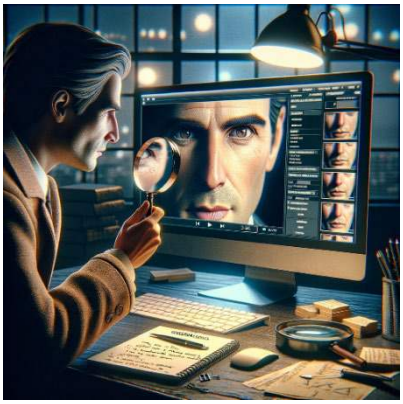
Semne de identificare a unui Deepfake

Protejarea împotriva Deepfake-urilor este o provocare continuă, deoarece tehnologia se perfecționează constant, făcând conținutul manipulat tot mai greu de distins de cel real. Cu toate acestea, există anumite indicii care pot trăda un Deepfake, iată la ce anume trebuie să fim atenți:

- Mediul înconjurător (de exemplu, umbre inexistente, reflexii prea puternice, zone neclare)
- Imperfecțiuni ale feței (alunițe nerealiste, clipire nesincronizată, distorsiuni în interiorul gurii cum ar fi lipsa dinților și a limbii, dinți mult prea perfecți etc.)
- Nesincronizarea vorbirii/sunetului și a mișcării buzelor, de exemplu: din cauza strănutului

Nesincronizarea vorbirii/sunetului și a mișcării buzelor poate fi observată la pronunțarea literelor *b, m* și *p*. Uneori apar pixeli în nuanțe de gri la marginile componentelor modificate. Se poate distinge dacă este vorba de o falsificare și atunci când persoana din înregistrare este privită dintr-un alt unghi. Dacă pentru crearea conținutului Deepfake nu s-au folosit fotografii ale persoanei din unghiuri diferite, algoritmul nu poate deduce aspectul persoanei din alt unghi, rezultând distorsiuni.

Massachusetts Institute of Technology (MIT) este una dintre entitățile care dezvoltă instrumente ML care pot identifica dacă un conținut este autentic sau un Deepfake, oferind totodată un chestionar educativ pentru a învăța publicul cum să facă această distincție[iv].



Un detector clasic de conținut vizual fals se bazează pe detectarea erorilor rezultate din prelucrare. Cel mai adesea, aceasta implică analiza pixelilor pe care ochiul uman nu îi poate vedea, deoarece prin manipularea imaginii marginile componentei modificate au caracteristici speciale. Un astfel de algoritm este hibridul dintre **Long short term memory** (LTSM – o rețea neuronală) și algoritmul **Encoder-Decoder**. Acesta funcționează analizând în paralel fiecare pixel individual și/sau întreaga imagine/videoclip comprimat(ă). În cele din urmă, rezultatele celor două funcții sunt comparate și dacă ambele indică aceeași regiune, materialul este considerat modificat.

Există, de asemenea, diferite tipuri de detectoare de acces. Detectorul Intel[v] se bazează pe observarea unor indicii subtile, invizibile ochiului uman, pentru a verifica autenticitatea conținutului. Prin analiza semnelor precum dilatarea pupilelor, modificări ale culorii vaselor de sânge în concordanță cu bătăile inimii etc., se poate determina dacă acel conținut este autentic[vi].

Instrumente și tehnici de detectare a Deepfake

Detectarea Deepfake-urilor este o provocare în continuă evoluție, întrucât tehnologiile de IA care stau la baza creării Deepfake-urilor devin din ce în ce mai sofisticate. Ca răspuns, cercetătorii și dezvoltatorii lucrează la dezvoltarea de instrumente și tehnici noi pentru a identifica aceste falsuri. Iată câteva dintre cele mai promițătoare abordări în detectarea Deepfake-urilor:

Analiza comportamentală: Această metodă se bazează pe identificarea micilor imperfecțiuni sau anomalii în comportamentul sau mișcările fizice ale subiectului din videoclip.

Consistența iluminării: Detectarea inconsistențelor în iluminare este o altă tehnică eficientă. Algoritmii analizează umbrele, reflexiile și modul în care lumina se reflectă pe diferite suprafețe ale feței pentru a determina dacă imaginea a fost manipulată.

Analiza texturii pielii: Tehnicile de Deepfake adesea netezesc textura pielii sau introduc anomalii la nivelul texturii. Detectarea acestor modificări, care sunt adesea subtile și greu de observat cu ochiul liber, poate ajuta la identificarea manipulărilor.

Detectarea artefactelor de compresie: Videoclipurile și imaginile manipulare prin IA pot prezenta artefacte de compresie unice datorită procesului de generare și compresie. Analiza acestor artefacte poate oferi indicii că materialul a fost alterat.

Examinarea metadatelor: Deși Deepfake-urile în sine pot fi convingătoare, metadatele asociate cu un fișier video sau imagine (cum ar fi data creării, tipul camerei etc.) pot fi contradictorii sau suspecte, sugerând manipulare.

Verificarea consistenței respirației și a pulsului: Unele tehnici avansate includ analiza variațiilor minore în culoarea feței sau a umbrelor, care pot indica bătăile inimii și respirația. Modificările în aceste modele pot indica prezența unui Deepfake.

Utilizarea rețelelor neuronale convoluționale: CNN-urile sunt folosite pentru a analiza videoclipurile frame cu frame, învățând caracteristicile specifice videoclipurilor autentice comparativ cu cele false. Această metodă poate fi extrem de eficientă, dar necesită o cantitate mare de date de antrenament.

Aplicații și servicii comerciale: Există și instrumente comerciale disponibile sau soluții oferite de start-up-uri specializate în detectarea Deepfake-urilor, care combină tehnologiile menționate mai sus pentru a oferi soluții la cheie pentru organizații sau indivizi.

Provocări și limitări: Este important de menționat că, pe măsură ce tehnologiile de detectare devin mai sofisticate, la fel devin și metodele de creare a Deepfake-urilor. Asta înseamnă că este nevoie de o actualizare și îmbunătățire constantă a instrumentelor de detectare. În plus, multe dintre tehnici pot genera răspunsuri fals pozitive sau fals negative, ceea ce necesită verificări suplimentare și îmbunătățirea continuă a algoritmilor.

Detectarea Deepfake-urilor este un câmp de bătălie dinamic între creatorii de conținut falsificat și cei care încearcă să protejeze autenticitatea informațiilor. Continuarea cercetării și dezvoltării în acest domeniu este crucială pentru a ține pasul cu evoluția rapidă a tehnologiei.

Sfaturi pentru a evita să fii păcălit de Deepfake-uri

Evitarea înșelăciunii prin Deepfake necesită o combinație de **scepticism sănătos, atenție la detalii și utilizarea unor instrumente de verificare**. Iată câteva sfaturi utile:

- Nu crede tot ce vezi online! Internetul este o sursă vastă de informații, dar nu toate sunt veridice. Este important să dezvolți un scepticism sănătos și să analizezi cu atenție orice conținut video sau foto înainte de a-l accepta ca fiind real.
- Caută semne de manipulare: Deepfake-urile pot fi foarte sofisticate, dar adesea pot fi identificate prin anumite indicii. Fii atent la discrepanțe de iluminare, erori de aliniere, nereguli ale pielii sau probleme de sincronizare a buzelor cu sunetul.

- Verifică sursa: De unde provine videoclipul sau imaginea? Este distribuit pe o platformă de încredere? Caută confirmarea informației din surse credibile sau direct de la entitățile sau persoanele implicate.
- Folosește instrumente de verificare: Există numeroase organizații și instrumente online care te pot ajuta să verifici dacă o informație este reală. Utilizează-le pentru a cerceta autenticitatea conținutului suspect.
- Nu te baza pe o singură sursă: Caută confirmare din mai multe surse credibile. Un singur videoclip sau imagine nu este suficient pentru a verifica o informație.
- Învață despre Deepfake-uri: Cu cât înțelegi mai bine cum funcționează această tehnologie, cu atât vei fi mai capabil să identifici falsurile. Există multe resurse online care explică principiile Deepfake-urilor și metodele de detectare.

Prin aplicarea acestor sfaturi, poți reduce riscul de a fi înșelat de conținutul Deepfake și poți contribui la promovarea unei culturi a verificării și a responsabilității în mediul online.

Educația în combaterea Deepfake-urilor

Informarea publicului: Educația este esențială pentru a crește gradul de conștientizare cu privire la Deepfake-uri și la pericolele asociate. Oamenii trebuie informați despre modul în care Deepfake-urile pot fi utilizate pentru manipulare și dezinformare, pentru a reduce impactul lor negativ.

Dezvoltarea gândirii critice: Programele educaționale pot contribui la dezvoltarea abilităților de gândire critică. Oamenii trebuie învățați cum să evalueze sursele de informații, să recunoască semnele de conținut falsificat și să verifice informațiile înainte de a le distribui.

Securitate digitală: Educația privind securitatea online și confidențialitatea poate ajuta indivizii să își protejeze mai bine propriile informații și să fie conștienți de potențialele abuzuri ale tehnologiei Deepfake.

Alfabetizare media: Înțelegerea modului în care funcționează mass-media și a tehnicilor de producție a conținutului poate ajuta publicul să discearnă mai bine Deepfake-urile.

Combinația dintre eforturi tehnologice și programe educaționale solide poate construi un răspuns puternic și eficient la provocările prezentate de Deepfake-uri. Această abordare dual este esențială pentru a minimiza impactul negativ al Deepfake-urilor asupra societății, politicilor și vieții private a cetățenilor.

Ce să faci dacă ești victima unui Deepfake

Dacă te afli în situația neplăcută de a fi victima unui Deepfake, este important să acționezi rapid și eficient pentru a minimiza daunele. Iată câțiva pași pe care îi poți urma:

- **Documentează abuzul**: Salvează copii ale conținutului Deepfake, inclusiv URL-uri, capturi de ecran, sau orice altă formă de dovezi care ar putea fi relevante. Acest lucru este esențial pentru orice demersuri legale sau raportări ulterioare.
- **Raportează conținutul**: Majoritatea platformelor de socializare și site-urilor web au politici stricte împotriva Deepfake-urilor și a conținutului manipulat. Utilizatorii pot semnala cu ușurință conținutul suspect folosind funcția de raportare integrată a platformei.
- **Contactează autoritățile**: În cazuri grave, unde conținutul Deepfake încalcă legile privind defăimarea, hărțuirea sau distribuția de materiale pornografice fără consimțământ, poate fi necesar să contactezi autoritățile locale sau alte organisme de aplicare a legii.
- **Solicită ajutor juridic**: Consultă un avocat pentru a evalua opțiunile legale disponibile pentru tine. Aceasta poate include acțiuni legale împotriva celor care au creat sau distribuit conținutul Deepfake.
- **Folosește servicii de gestionare a reputației online**: Există companii specializate în îmbunătățirea prezenței online și în eliminarea sau diminuarea impactului conținutului negativ. Aceste servicii pot fi utile pentru a-ți proteja imaginea pe termen lung.
- **Comunică cu transparență**: Dacă Deepfake-ul are potențialul de a-ți afecta cariera sau relațiile personale, ia în considerare să vorbești deschis despre situație cu angajatorul, colegii sau persoanele apropiate. Oferirea contextului și a versiunii tale poate ajuta la diminuarea impactului negativ.
- **Protejează-ți informațiile personale**: În urma unui incident Deepfake, este important să fii mai precaut în ceea ce privește securitatea online. Verifică setările de confidențialitate pe rețelele sociale, schimbă parolele și monitorizează activitatea conturilor tale pentru semne de acces neautorizat.
- **Suport emoțional**: Trauma psihologică suferită de victima unui Deepfake poate fi semnificativă. Nu ezita să cauți sprijin din partea prietenilor, familiei sau a profesioniștilor în sănătate mintală.
- **Educație și conștientizare**: Ajută la creșterea gradului de conștientizare despre pericolele Deepfake-urilor prin împărtășirea experienței tale, dacă te simți confortabil. Acest lucru poate ajuta la informarea și protejarea altora.

Acționând prompt și decisiv, poți trece prin provocările asociate cu a fi victima unui Deepfake și poți începe procesul de recuperare și protecție a reputației tale.

Regulament privind inteligența artificială adoptată de Parlamentul European

Parlamentul European a adoptat Regulamentul COM/2021/206 privind IA care are ca scop protejarea drepturilor fundamentale, democrației, statului de drept și sustenabilității mediului. Regulamentul clasifică sistemele IA în funcție de nivelul de risc (**inacceptabil, ridicat, limitat, minim**) și interzice anumite utilizări, cum ar fi manipularea cognitiv-comportamentală a persoanelor vulnerabile, sistemele de credit social și identificarea biometrică în timp real (cu excepții). De asemenea, regulamentul impune obligații de transparență, trasabilitate, evaluare a riscurilor și asigurare a calității.

Regulamentul abordează și problema Deepfake-urilor, definindu-le ca tehnici de manipulare a imaginilor sau videoclipurilor pentru a crea iluzia că un subiect spune sau întreprinde o acțiune ce nu a avut loc sau nu în contextul prezentat. Deepfake-urile cu risc inacceptabil sunt interzise, cele cu risc ridicat trebuie să

fie identificabile ca atare, iar cele cu risc limitat și minim nu sunt supuse unor restricții specifice. Furnizorii de Deepfake-uri trebuie să se asigure că utilizatorii sunt conștienți de natura artificială a conținutului, iar platformele online trebuie să ia măsuri pentru a preveni răspândirea Deepfake-urilor dăunătoare[vii].

Concluzii

Pe măsură ce tehnologia avansează, capacitatea de a crea Deepfake-uri devine tot mai accesibilă, sporind necesitatea unor strategii eficiente de detecție și conștientizare.

Dezvoltarea algoritmilor de detectare: Cercetătorii lucrează la dezvoltarea de algoritmi bazați pe IA care pot identifica Deepfake-urile prin analizarea detaliilor care sunt dificil de detectat de ochiul uman, cum ar fi anomalii în clipire sau în mișcarea naturală a pielii.

Tehnici de autentificare a conținutului: Tehnologii precum blockchain și watermarking digital (filigranare) pot ajuta la stabilirea originii și autenticității conținutului, oferind o metodă de verificare a surselor video și audio originale.

Instrumente de verificare la scară: Platformele de social media și companiile tehnologice dezvoltă instrumente automatizate pentru a scana și a elimina Deepfake-urile de pe site-urile lor, contribuind la reducerea răspândirii dezinformării.

Parteneriate între industrii: Colaborarea între sectorul tehnologic, guvern și organizațiile de cercetare poate accelera dezvoltarea și implementarea soluțiilor de detectare a Deepfake-urilor.

Acest material a fost realizat de următorii experți ai DNSC:

Daniel Abotezătoaei, Dan Andrieș, Mihaela Dan, Alex Leoreanu, Irina Nemoianu, Cristian Nistor, Vlad Drăguș, Mihai Rotariu

i https://www.digi24.ro/stiri/cum-a-fost-lasata-o-pensionara-fara-52-de-mii-de-lei-am-intrat-pe-youtube-unde-ascult-rugaciuni-si-am-dat-de-un-videoclip-cu-tiriac-2714127

ii https://consumer.ftc.gov/consumer-alerts/2023/03/scammers-use-ai-enhance-their-family-emergency-schemes

iii https://www.forbes.com/sites/jessedamiani/2019/09/03/a-voice-deepfake-was-used-to-scam-a-ceo-out-of-243000/

iv https://detectfakes.kellogg.northwestern.edu/

v https://www.intel.com/content/www/us/en/newsroom/news/intel-introduces-real-time-deepfake-detector.html

vi https://www.cert.hr/wp-content/uploads/2023/12/zloupotreba_umjetne_inteligencije.pdf

vii https://www.europarl.europa.eu/news/ro/press-room/20240308IPR19015/legea-privind-inteligenta-artificiala-pe-adopta-un-act-de-referinta